

APPLICATION FOR UNITED STATES LETTERS PATENT

For

**A METHOD FOR DYNAMIC ASSIGNMENT OF SLOT-DEPENDENT
STATIC PORT ADDRESSES**

Inventors:

Kuriappan P. Alappat
Chetan Hiremath

Prepared by:

BLAKELY SOKOLOFF TAYLOR & ZAFMAN LLP
12400 Wilshire Boulevard
Los Angeles, CA 90025-1026
(206) 292-8600

Attorney's Docket No.: 42.P17958

"Express Mail" mailing label number: EV320120270US

Date of Deposit: December 3, 2003

I hereby certify that I am causing this paper or fee to be deposited with the United States Postal Service "Express Mail Post Office to Addressee" service on the date indicated above and that this paper or fee has been addressed to the Mail Stop New Application, Commissioner for Patents, P.O. Box 1450, Alexandria, VA 22313-1450

Christina Fernandez

(Typed or printed name of person mailing paper or fee)

Christina Fernandez

(Signature of person mailing paper or fee)

December 3, 2003

(DATE SIGNED)

A METHOD FOR DYNAMIC ASSIGNMENT OF SLOT-DEPENDENT STATIC PORT ADDRESSES

FIELD OF THE INVENTION

[0001] The field of invention relates generally to modular platforms and, more specifically but not exclusively relates to a method for dynamically assigning slot-dependent static port addresses.

BACKGROUND INFORMATION

[0002] In a typical card modular platform, a plurality of boards are installed in a shelf having a backplane to which each board is communicatively coupled. For example, Figure 1 shows a shelf 100 having an integrated subrack 102 including a plurality of slots into which respective boards 104 are installed. At the rear end of the integrated subrack 102 is a backplane 106. The backplane enables inter-board communication via a well-known communication scheme, such as Ethernet, Fibre Channel, PCI ExpressAS, etc. The communication links may also be TDM (Time Division Multiplexed) fabrics, such as Sonet/SDH or buses such as PCI. Under any of these schemes, each board is identified by a unique slot address.

[0003] In addition, boards may have common IP addresses in certain configurations. For example, if there is an active and standard by board, it is possible for a common IP address to follow the active board. In case of boards in multiple slots do the same function in a load sharing environment, it is possible to have a common IP address for all those boards where a switch implements load sharing using one of many load sharing algorithms. In that case, the traffic is routed through the appropriate ports to different slots.

[0004] One common approach is to employ Ethernet for inter-board communication, as illustrated in Figure 2, which shows the internal configuration of a telecommunications (telco) shelf 200. Switch 200 includes a dual star Ethernet

fabric configured with a pair of switch fabrics 202A and 202B, each of which is coupled in communication to a plurality of boards 204A-G via a common backplane (not shown). The two switch fabrics 202A and 202B route traffic between any two boards, while the dual star configuration provides 1 + 1 redundancy. Switch
5 fabrics 202A and 202B are also coupled in communication with a redundant external core switch 206, which provides an interface to an external network 208.

- [0005] In general, telco equipment and the like need to provide 24-7 availability, thus the use of redundancy in the switch fabric and external core switch. At the same time, each slot within a given shelf will be configured for a particular function.
10 Once configured, the configuration remains the same unless a major event warrants reconfiguration. Redundant boards may be used to provide non-stop service. In case of a failure, automatic fail-over to the standby board takes place and the failed board is eventually replaced. This sequence ensures that the operation is not interrupted.
15 [0006] Among the configuration operations is the assignment of an IP address for each of a board's Ethernet ports. In a typical installation, each slot within a shelf may be configured for a particular function, with the possibility of multiple slots used to perform the same function in a load sharing manner. It is possible that boards in certain slots are loaded with pre-configured set of software, including operating
20 systems, drivers, applications, configuration settings, etc. Furthermore, at least a portion of the boards, such as the system manager board/s, is aware of the functions served by the other boards. Accordingly, it is advantageous to assign an IP address for a replacement board that matches the IP address of the board being replaced. Under this consideration, each board is assigned a static IP address, as illustrated
25 by IP addresses 210, as illustrated in Figure 2. Also it is advantageous to load software images dependent on slot address since as described earlier each slot is pre-assigned a particular function. This allows for ease of deployment in a modular

platform environment by avoiding manual intervention, potential human errors and saving time.

[0007] There are several conventional methods for assigning static IP addresses.

One method is to manually configure the board with a pre-assigned address. This is

5 time consuming and error prone. It also requires the board to be up and running

prior to assigning an address. Another scheme is to employ a DHCP (Dynamic Host

Configuration Protocol) server, which is normally used to dynamically assign

addresses. However, in this instance, it is desired to assign the new replacement

board a pre-determined address. DHCP supports this option via an address

10 reservation mechanism. The mechanism works by knowing the MAC (media access

channel) address of the board's network port. The MAC address is entered into a

DHCP console or the like, along with the pre-assigned address. When the board is

installed, a DHCP message exchange occurs during which the board identifies its

MAC address and the DHCP server issues the corresponding pre-assigned address.

15 In a somewhat similar manner, IP addresses may be assigned by an Ethernet

switch.

[0008] While the foregoing DHCP scheme works for assigning static IP

addresses, it has its limitations. For example, a critical requirement for all networks

is each node must have a unique network address. Most, if not all, DHCP server

20 configuration schemes do not let the same IP address be reserved for more than

one MAC address, and thus one board. As a result, it is not possible to configure IP

addresses for replacement boards in advance using a conventional DHCP

reservation. It is also more complex to access the DHCP server configuration and

make changes as the access to it is limited and could consume more time. Besides,

25 the administrator who replaces the board may not have access privileges to make

changes to network configurations.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] The foregoing aspects and many of the attendant advantages of this invention will become more readily appreciated as the same becomes better understood by reference to the following detailed description, when taken in conjunction with the accompanying drawings, wherein like reference numerals refer to like parts throughout the various views unless otherwise specified:

[0010] Figure 1 is a schematic diagram illustrating a typical configuration for a card modular platform including a plurality of boards inserted into respective slots in an integrated sub-rack in a shelf;

10 [0011] Figure 2 is a schematic diagram illustrating various components and intercommunication paths between various boards in a telecommunications shelf configured as a card modular platform;

15 [0012] Figure 3 is a flow diagram illustrating interactions between and operations performed by a client board, a DHCP server, and a PXE server during a static network address assignment process, according to one embodiment;

[0013] Figure 4 is a flowchart illustrating operations performed via execution of instructions on a card modular platform board to obtain a static network address for the board; and

20 [0014] Figure 5 is a schematic diagram illustrating various components of an exemplary card modular platform board that may be used to practice embodiments of the invention disclosed herein.

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

[0015] Embodiments of methods and apparatus for dynamic assignment of slot-dependent static port addresses and software images are described herein. In the following description, numerous specific details are set forth, such as embodiments in which IP network addresses are assigned, to provide a thorough understanding of embodiments of the invention. One skilled in the relevant art will recognize, however, that the invention can be practiced without one or more of the specific details, or with other methods, components, materials, etc. In other instances, well-known structures, materials, or operations are not shown or described in detail to avoid obscuring aspects of the invention.

[0016] Reference throughout this specification to "one embodiment" or "an embodiment" means that a particular feature, structure, or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, the appearances of the phrases "in one embodiment" or "in an embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures, or characteristics may be combined in any suitable manner in one or more embodiments.

[0017] In accordance with aspects of this specification, techniques are disclosed for automatically assigning static IP addresses based on the shelf and slot addresses of respective boards in a given card modular equipment where the equipment could be a single shelf or multiple shelves in a rack. In case of multiple shelves, each slot in each shelf will have a unique slot address. Under the technique, a board does not depend on either a DHCP server, fabric board, or manual intervention for the assignment of an static network port IP address (or addresses for a multi-port board). Rather, a static network address is assigned in a

manner that is independent of the MAC address of the board, eliminating the overhead and errors associated with the prior art schemes.

[0018] An exemplary technique for obtaining static network port addresses in accordance with one embodiment is depicted in Figure 3. The technique involves a series of message exchanges between a client board 300, a DHCP server 302, and a PXE (pre-boot execution environment) server 304. In one embodiment, the DHCP server and PXE server are co-located on the same machine. The series of message exchanges illustrated in Figure 3 correspond to operations that are performed in response to a board start-up event.

10 [0019] First, a small portion of firmware 306 is executed on client board 300 to perform early initialization of the board, including enabling basic network communications. For example, to support Ethernet networks, a network interface that supports Ethernet is initialized.

15 [0020] The next set of operations involves an exchange of messages between client board 300 and DHCP server 302 to obtain a temporary IP address. For simplicity, this message exchange is depicted as a DHCP IP address request and a DHCP acknowledge. In practice, the series of communications exchanges comprises the following:

1. The client board broadcasts a DHCP_Discover message on its local sub-net searching for DHCP server; the request may go over subnet boundaries if the switches are set up to relay the requests
2. A listening DHCP server sends a DHCP_Offer message containing an offered IP address to the client board;
3. The client board accepts the offered IP address and broadcasts a DHCP_Request message on the local sub-net containing the accepted IP address; and

4. The DHCP server responds via a unicast to the client board with a
DHCP_Ack message to acknowledge the IP address has been accepted.

[0021] The foregoing illustrates a sequence under which a single DHCP server receives the DHCP_Discover message. Under some circumstances, multiple DHCP servers may receive the DHCP_Discover message, and offer respective IP addresses in response. Under this circumstance, the client board will select one of the offered IP addresses. The net result is that the client board will end up with a temporary IP address 308. The particular address is not important, and will generally relate to the IP address scope allotted to the DHCP server by an administrator. At this point, the client board 300 can communicate with other network entities via unicasts rather than broadcasts.

[0022] The remaining message exchanges are between the client board 300 and the PXE server 304 (or a co-located DHCP/PXE server). In general, a PXE server is used to provide bootable operating system (OS) images to network clients, thus removing the requirement of the client needing to store a local OS image and applications on local hard disk drives (HDDs) or flash ROMs (read-only memories). Even if images and applications are stored locally in flash memory or on a local HDD, the same technique may be used to update the OS and image. In addition to this function, PXE server 304 also serves the function of a network address proxy.

20 That is, the PXE server is configured to allocate network address in lieu of a conventional address allocated, such as a DHCP server or a domain controller.

[0023] In order to exchange messages with PXE server 304, client board 300 needs to know the PXE server's network address, and a transmission protocol needs to be established. In one embodiment, if the DHCP and PXE servers reside 25 on the same machine, the response to the DHCP request above will contain information needed by the client to start a TFTP (Trivial File Transfer Protocol) session. TFTP is a simplified transmission protocol that does not require the

overhead of more robust protocols, such as the TCP/IP protocol used for most network traffic. If the DHCP and PXE servers are hosted by separate machines (necessitating separate network addresses) and the DHCP server is configured to know the IP address of PXE server 304, the PXE server's address will be included in
5 the DHCP message exchange, such as depicted by a PXE server address 309. The client board 300 will then contact PXE server 304 via PXE server address 309 to obtain information for starting a TFTP session. If the DHCP server does not have address information for the PXE server, the client may broadcast a PXE discover message akin to the DHCP discover message discussed above to locate the PXE
10 server. Upon receiving the PXE discover message, the PXE server will respond with information for starting a TFTP session, including it's network address.

[0024] The series of message exchanges between client board 300 and PXE server 3000 begins with a PXE download request message sent from the client board to the PXE server via TFTP. In response to receiving the request, the PXE
15 server 304 returns an initial boot image 310 via TFTP to the client board.

[0025] Upon receiving the initial boot image 310, the image is executed by the client board, as depicted by a block 312. In general, the initial boot image comprises a generic or common boot image that is applicable for initializing various different types of client boards and is independent of a board's slot number. Included in the
20 initial boot image (or otherwise already stored on the client board) are instructions to enable the client board to obtain its shelf and slot addresses in a block 314. This operation may be performed in one of several manners, depending on the particular board type and/or configuration.

[0026] For example, in one embodiment client board 300 is configured as a
25 PICMG 3.0 (PCI (peripheral component interface) Industrial Computers Manufacturing Group)-compliant board. Accordingly, client board 300 sends IPMI (Intelligent Platform Management Interface) commands *GetAddressInfo* and

GetShelfAddressInfo over a KCS (Keyboard Controller Style, an IPMI messaging protocol) interface to obtain the slot and shelf addresses, respectively. If a KCS interface is not implemented, similar information may be retrieved through another type of interface, such as a propriety interface.

- 5 [0027] In further detail, a *GetAddressInfo* command is sent by the initial boot image on the client board to its own IPMC (IPMI Controller). The IPMC's response contains several fields including one called "Site ID" which is the same as the board's slot number. The response also contains its IPMB (Intelligent Peripheral Management Bus) address, which is used in the "Send Message" command
10 described below. The request and response data are described in Table 3-8, of the PICMG 3.0 specification.

- [0028] Next, a *GetShelfAddressInfo* command (as per Table 3-12, PICMG 3.0) is sent by the initial boot image to the Shelf Manager (at 20h IPMB address) using the "Send Message" command format described in Table 18-9, IPMI V1.5 Specification.
15 The initial boot image on the client board then sends a *GetShelfAddressInfo* command to its IPMC to fetch response data. The response contains the Shelf Address field.

- [0029] In another embodiment, client board 300 is configured in accordance with the CompactPCI standard. In this case, the board's OEM (original equipment manufacturer) may define OEM-specific IPMI commands to determine the slot and shelf address to uniquely identify the board in a given facility.
20

- [0030] In a block 316, a final boot image is determined for the client board. Unlike the generic initial boot image 310, the final boot image will typically pertain to a particular board configuration or type. In one embodiment, a set of available final
25 boot images is provided to the client board in the initial boot image 310. After the final boot image has been determined, a download request for the same is sent via

TFTP from the client board to PXE server 304. The TFTP request also includes slot and shelf address data 318.

[0031] In response to receiving the final boot image download request, PXE server 304 retrieves a corresponding image and returns it to client board 300. At the 5 same time, the PXE server determines an IP address (or multiple addresses in the case of multi-port boards) to allocate client board 300 based on shelf and slot data 318. In one embodiment, a pre-configured lookup table 320 is stored by PXE server 304, or otherwise made available for access to the PXE server. For example, in the latter instance a central configuration server may store an IP address lookup 10 table for an entire system that uses multiple PXE servers.

[0032] An exemplary configuration for lookup table 320 is shown in Figure 3. The illustrated configuration includes a shelf column, a slot column, and an address column. The address to be assigned is determined based on the row containing shelf and slot values corresponding to shelf and slot data 318.

[0033] In one embodiment, a final boot image 322 is returned to the client 15 board 300, along with an embedded or attached assigned IP address 324. In another embodiment, the final boot image and assigned IP address are sent as separate TFTP payloads.

[0034] The address configuration process is completed in a block 326, wherein 20 the final boot image 322 is loaded and/or executed, and a static IP address corresponding to assigned IP address 324 is assigned by the operating system of client board 300. Henceforth, other system components, such as boards 204A-G, will be able to communicate with the new board via the assigned IP address 324.

[0035] Figure 4 illustrates operations performed during an embodiment of a static 25 network address allocation scheme that is applicable for boards that boot an operating system stored on a local device. The process begins with a start block 400 corresponding to a power-on or reset event. For example, inserting a new

board into a slot would result in a power-on event. In response, a series of pre-boot system initialization operations are performed in a block 402. This generally encompasses the loading and execution of firmware to set up the board for booting an operating system. For example, in response to a power-on event, this will 5 typically include performing a Power-On Self Test (POST), as well as memory initialization. For reset events, the POST operations are usually skipped.

[0036] Next, determination of the boot source is performed, as depicted by a block 404 and decision block 406. For example, a bootable image may be loaded from a flash device, a hard disk (HD), a CD-ROM, a floppy disk (FD), etc. Once the 10 boot source is determined, the bootable OS image is booted (i.e., loaded and/or executed) in a block 408.

[0037] The remaining operations in blocks 410, 412, and 414 relate to obtaining a static network address. In one embodiment, the bootable OS image includes instructions for performing these operations. In another embodiment, firmware 15 stored on the client board includes instructions for performing the operations. In yet another embodiment, the instructions are provided via a combination of firmware and the operating system image.

[0038] In block 410 the shelf and slot addresses are obtained. This process is similar to that discussed above for the PXE Server embodiment. For example, 20 *GetAddressInfo* and *GetShelfAddressInfo* IPMI commands may be employed for PICMG-compliant boards, or OEM IPMI commands may be used for configurations such as CompactPCI-based platforms.

[0039] Next, in block 412, an IP address is determined using the shelf and slot addresses as inputs. In one embodiment, an algorithm is employed to generate the 25 addresses. For installations in which local booting is used system-wide, a single algorithm may be employed to guarantee unique IP address assignment. For example, the shelf address may be used to generate an IP sub-mask, while the slot

address may be used to generate the non-masked portion of the IP address. In another embodiment, an IP address lookup table similar to lookup table 320 may be used. In general, a copy of the IP address lookup table may be stored on the client board, locally accessible to the client board (e.g., stored in a component accessible to all boards within a common shelf), or accessible via a network connection. In this latter case, means for accessing the network will be enabled prior to retrieving the lookup table.

[0040] The process is completed in a block 414, wherein the network address determined in block 412 is assigned as a static IP address by the operating system. 10 This operation can be performed using well-known techniques particular to a given operating system; accordingly, further details are not provided herein.

[0041] The foregoing schemes provide several advantages over known network address allocation schemes. It enables field-replaceable units to readily identify itself in highly dense modular server environments. The network address of a failed 15 board and its replacement are the same, enabling the replacement process to appear completely transparent to other system components. There is no manual configuration required, nor is there any special programming that needs to be made at a DHCP server or Ethernet switch. The scheme is easily scalable, with lookup tables being support both at a local PXE server and centralized server levels.

[0042] Figure 5 shows a board 500 suitable for practicing client-side and stand-alone aspects of the embodiments disclosed herein. Board 500 includes a printed circuit board (PCB) 502 on which various components are operatively coupled, including one or more processors 504, and memory 506. Memory 506 may include, but not limited to, Dynamic Random Access Memory (DRAM), Static Random 25 Access Memory (SRAM), Synchronized Dynamic Random Access Memory (SDRAM), Rambus Dynamic Random Access Memory (RDRAM), or the like. For the purpose of clarity, details of various well-known circuit elements, such as power

planes, busses, and the like are not shown on PCB 502; it will be understood that such circuit elements would be present in an actual board. Likewise, various integrated circuit and passive circuit elements (e.g., resistors, capacitors, inductors, etc.) are also not shown for clarity. For purpose of illustration, the various 5 components depicted in Figure 5 are linked in communication via a bus 507, which will be understood to represent various types of busses that may be provided by PCB 502, including but not limited to parallel, serial, PCI, compactPCI, etc.

[0043] Firmware for the board and/or persistent data will typically be stored on some type of non-volatile storage device 508. Non-volatile storage devices include, 10 but are not limited to, Read-Only Memory (ROM), Flash memory, Erasable Programmable Read Only Memory (EPROM), Electronically Erasable Programmable Read Only Memory (EEPROM), or the like memory. In cases where OS and image sizes are large, for example, IA (Intel Architecture) single board computers (SBCs), the images are stored in volatile memory such as DRAM. In this 15 case, the image is loaded into the volatile memory every time power is cycled.

[0044] Each board will include a set of connectors 510 that are used to couple the board to a common backplane or the like. In general, the set of connectors 510 will be compatible with the backboard standard or configuration employed for the platform, and are used to provide inter-board communication paths via the 20 backplane, as well as communication paths to external network entities. The backplane (not shown) will typically provide power to each board (e.g., -48 volts for telco boards), although a power converter such as a DC-DC converter may be provided by a board to ensure appropriate voltage and current levels are available for the various components mounted to PCB 502. A network interface 512 will be 25 provided to facilitate network communication, such as but not limited to an Ethernet network interface controller.

[0045] Some boards may provide various types of mass storage, including a hard disk drive 514. Optional mass storage devices that may also be provided include removable machine-readable media peripherals, such as a CD-ROM drive 516, floppy drive 518, Zip drive, tape drive 520, etc.

5 [0046] The machine-executable instructions for performing the various board operations discussed above will generally be embodied as firmware and/or software instructions that are executed on processor(s) 504. The firmware and/or software instructions will typically be provided via a machine-readable medium. For the purposes of this specification, a machine-readable medium includes any mechanism 10 that provides (i.e., stores and/or transmits) information in a form readable or accessible by a machine (e.g., one or more processors 504). For example, a machine-readable medium includes, but is not limited to, recordable/non-recordable media (e.g., a read only memory (ROM), a random access memory (RAM), a magnetic disk storage media, an optical storage media, a flash memory device, 15 etc.). In addition, a machine-readable medium can include propagated signals such as electrical, optical, acoustical or other form of propagated signals (e.g., carrier waves, infrared signals, digital signals, etc.).

[0047] The above description of illustrated embodiments of the invention, including what is described in the Abstract, is not intended to be exhaustive or to 20 limit the invention to the precise forms disclosed. While specific embodiments of, and examples for, the invention are described herein for illustrative purposes, various equivalent modifications are possible within the scope of the invention, as those skilled in the relevant art will recognize.

[0048] These modifications can be made to the invention in light of the above 25 detailed description. The terms used in the following claims should not be construed to limit the invention to the specific embodiments disclosed in the specification and the claims. Rather, the scope of the invention is to be determined entirely by the

following claims, which are to be construed in accordance with established doctrines of claim interpretation.